# TRANSMITTAL OF APPEAL BRIEF (Large Entity)

Docket No.
US010324/16943

In Re Application Of:   Eric Cohen-Solal

| Serial No. 09/896,199 | Filing Date June 29, 2001 | Examiner Paul A. Bell | Group Art Unit 2675 |
|---|---|---|---|

Invention:   **PICTURE-IN-PICTURE REPOSITIONING AND/OR RESIZING BASED ON SPEECH AND GESTURE CONTROL**

**RECEIVED**

MAR 3 1 2004

## TO THE COMMISSIONER FOR PATENTS:

Technology Center 2600

Transmitted herewith in triplicate is the Appeal Brief in this application, with respect to the Notice of Appeal filed on

The fee for filing this Appeal Brief is:     **$330.00**

☒ A check in the amount of the fee is enclosed.

☐ The Director has already been authorized to charge fees in this application to a Deposit Account.

☒ The Director is hereby authorized to charge any fees which may be required, or credit any overpayment to Deposit Account No.   **19-1013/SSMP**

_Signature_

Dated:   **March 24, 2004**

**Thomas Spinelli**
**Registration No.: 39,533**

**Scully, Scott, Murphy & Presser**
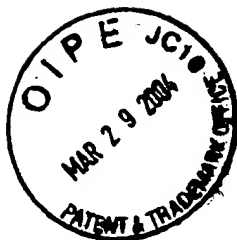**400 Garden City Plaza**
**Garden City, NY  11530**
**(516) 742-4343**

CC:

P30LARGE/REV03

#11 1/3

RECEIVED

MAR 3 1 2004

Technology Center 2600

**BRIEF ON APPEAL**

Thomas Spinelli
Attorney for Appellant
Registration No. 39,533

SCULLY SCOTT MURPHY & PRESSER
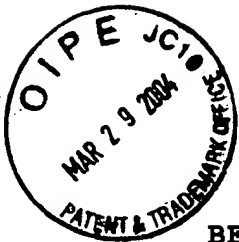400 Garden City Plaza
Garden City, New York 11530
(516) 742-4343

h\work\1472\16943\amend\16943.BR1

# TABLE OF CONTENTS

**IN THE UNITED STATES PATENT AND TRADEMARK OFFICE**

**BEFORE THE BOARD OF PATENT APPEALS AND INTERFERENCES**

| | |
|---|---|
| Applicant: Eric Cohen-Solal | Art Unit: 2675 |
| Serial No.: 09/896,199 | Examiner: Paul A. Bell |
| Filed: June 29, 2001 | Docket: US010324 (16943) |
| For: PICTURE-IN-PICTURE REPOSITIONING AND/OR RESIZING BASED ON SPEECH AND GESTURE CONTROL | Dated: March 24, 2004 |

Conf. No. 7568

Mail Stop Appeal Brief-Patents
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

**BRIEF ON APPEAL**

Sir:

## I. INTRODUCTION

Pursuant to the provisions of 35 U.S.C. §§ 134 and 37 C.F.R. §§ 1.191 and 1.192, this paper is submitted as a brief setting forth the authorities and arguments upon which Appellants rely in support of the appeal from the Final Rejection of Claims 1-20 in the above-identified patent application on November 5, 2003.

## II. REAL PARTY OF INTEREST

The real party of interest in the above-identified patent application is Koninklijke Philips Electronics N.V.

-1-

### III. RELATED APPEALS AND INTERFERENCES

Appellants respectfully submit that the present application is involved in no other appeal or interference besides the present Appeal.

### IV. STATUS OF THE CLAIMS

The parent application, U.S. patent application Serial No. 09/896,199 was filed on June 29, 2001, originally included Claims 1-18.

In an Official Action dated July 7, 2003, the Examiner rejected claims 1-18 under 35 U.S.C. § 103(a) as being unpatentable over U.S. Patent No. 5,999,214 to Inagaki (hereinafter "Inagaki") in view of Pavlovic et al., "Integration of Audio/Visual Information for Use in Human-Computer Intelligent Interaction," Image Processing, 1997 Proceedings IEEE, page 121 (hereinafter "Pavlovic").

In a Response under 37 C.F.R. § 1.111, filed August 18, 2003, with regard to independent claim 1, Applicant argued Inagaki does not disclose or suggest " a processor … configured … to change a PIP display characteristic in response to a received audio <u>command</u> and a related gesture from a user." Applicants further argued that since the system of Inagaki does not recognize and/or respond to spoken commands, there is no motivation or suggestion to combine it with the system of Pavlovic that is directed to recognition of spoken commands.

-2-

With regard to claim 11, the Applicant respectfully submitted that there is no motivation or suggestion to combine the references as discussed with regard to claim 1. The Applicant further submitted that neither of the cited references teach or suggest the recited features of claim 11.

With regard to claim 15, the Applicant respectfully submitted that there is no motivation or suggestion to combine the references as discussed with regard to claim 1. Furthermore, Applicant argued that claim 15 contains program segments for performing similar steps as recited in claim 11 and patentably distinguishes over the cited references for at least the same reasons as set forth with regard to claim 11.

Furthermore, the Applicant respectfully submitted that claims 2, 5, and 6 patentably distinguish over the cited references independently of their base claim (claim 1) and are allowable.

Lastly, new claims 19 and 20 were added.

A Final Official Action was issued on November 5, 2003, in which the Examiner rejected claims 1-20 as being unpatentable over Inagaki in view of Pavlovic. In response to the Final Official Action, Applicant substantially reiterated the arguments from the previous response.

Consequently, Claims 1-20 are the claims on appeal. A copy of the rejected claims is attached hereto in the Appendix.

### V.   **STATUS OF THE AMENDMENTS**

Appellants have not filed any amendments subsequent to the issuance of the Final Rejection of November 5, 2003.

### VI.   **SUMMARY OF THE INVENTION**

The present invention relates to a video display device comprising: a display configured to display a primary image and a picture-in-picture image (PIP) overlaying the primary image; and a processor operatively coupled to the display and configured to receive a first video data stream for the primary image, to receive a second video data stream for the PIP, and to change a PIP display characteristic in response to a received audio command and a related gesture from a user.

The present invention also relates to a method of controlling a display characteristic of a picture-in-picture display (PIP) overlaying a primary display. The method comprising: receiving an audio indication from a user; determining whether the received audio indication is one of a plurality of expected audio indications; analyzing a gesture of the user if the received audio indication is one of the plurality of expected audio indications; and controlling the

-4-

display characteristic if the gesture is a gesture related to the received audio indication.

The present invention also relates to a program segment stored on a processor readable medium for controlling a display characteristic of a picture-in-picture display (PIP) overlaying a primary display. The program segment comprising: a program segment for controlling receipt of an audio indication; a program segment for determining whether a received audio indication is one of a plurality of stored audio indications; a program segment for analyzing a gesture of the user if the received audio indication is one of the plurality of stored audio indications; and a program segment for controlling the display characteristic if the gesture is a gesture related to the received audio indication.

The present invention also relates to a video display device comprising: a display configured to display a primary image and a picture-in-picture image (PIP); and a processor operatively coupled to the display and configured to receive a first video data stream for the primary image, to receive a second video data stream for the PIP, and to change a PIP display characteristic in response to a received audio indication and a related gesture from a user, wherein the PIP display characteristic is at least one of a position of the PIP on the display and a display size of the PIP.

-5-

The present invention also relates to a video display device comprising: a display configured to display a primary image and a picture-in-picture image (PIP); and a processor operatively coupled to the display and configured to receive a first video data stream for the primary image, to receive a second video data stream for the PIP, and to change a PIP display characteristic in response to a received audio indication and a related gesture from a user, wherein the processor is configured to analyze image information received from the user after the audio indication is received to identify the change in the PIP display characteristic that is expressed by the received gesture.

The specification, from page 5 to page 16, discusses illustrative embodiments of the present invention in detail.

## VII.  **THE APPEALED CLAIMS**

Claims 1-20 are on appeal before the Board of Patent Appeals and Interferences, with Claims 1, 11, 15, 19, and 20 being the independent claims.  Independent Claim 1 is directed to a video display device comprising: a display configured to display a primary image and a picture-in-picture image (PIP) overlaying the primary image; and a processor operatively coupled to the display and configured to receive a first video data stream for the primary image, to receive a second video data stream for the PIP, and to change a PIP display

-6-

characteristic in response to a received audio command and a related gesture from a user.

Claims 2-10 directly or indirectly depend upon Claim 1 and further limit the scope of Claim 1.

Claim 11 is directed to a method of controlling a display characteristic of a picture-in-picture display (PIP) overlaying a primary display. The method comprising: receiving an audio indication from a user; determining whether the received audio indication is one of a plurality of expected audio indications; analyzing a gesture of the user if the received audio indication is one of the plurality of expected audio indications; and controlling the display characteristic if the gesture is a gesture related to the received audio indication.

Dependent Claims 12-14 directly or indirectly depend upon Claim 11 and further limit the scope of Claim 11.

Claim 15 is directed to a program segment stored on a processor readable medium for controlling a display characteristic of a picture-in-picture display (PIP) overlaying a primary display. The program segment comprising: a program segment for controlling receipt of an audio indication; a program segment for determining whether a received audio indication is one of a plurality of stored audio indications; a program segment for analyzing a gesture of the user if the

received audio indication is one of the plurality of stored audio indications; and a program segment for controlling the display characteristic if the gesture is a gesture related to the received audio indication.

Dependent Claims 16-18 directly or indirectly depend upon Claim 15 and further limit the scope of Claim 15.

Claim 19 is directed to a video display device comprising: a display configured to display a primary image and a picture-in-picture image (PIP); and a processor operatively coupled to the display and configured to receive a first video data stream for the primary image, to receive a second video data stream for the PIP, and to change a PIP display characteristic in response to a received audio indication and a related gesture from a user, wherein the PIP display characteristic is at least one of a position of the PIP on the display and a display size of the PIP.

Claim 20 is directed to a video display device comprising: a display configured to display a primary image and a picture-in-picture image (PIP); and a processor operatively coupled to the display and configured to receive a first video data stream for the primary image, to receive a second video data stream for the PIP, and to change a PIP display characteristic in response to a received audio indication and a related gesture from a user, wherein the processor is

-8-

configured to analyze image information received from the user
after the audio indication is received to identify the change
in the PIP display characteristic that is expressed by the
received gesture.

Each of the appealed claims, mentioned _supra_, is set
forth in the Appendix.

### VIII.  THE PRIOR ART RELIED UPON

The references relied upon by the Examiner in
rejecting Claims 1-20 are U.S. Patent No. 5,999,214 to Inagaki,
(hereinafter "Inagaki") and Pavlovic et al., "Integration of
Audio/Visual Information for Use in Human-Computer Intelligent
Interaction," Image Processing, 1997 Proceedings IEEE, page 121
(hereinafter "Pavlovic").

### IX.  THE ISSUE

The issue raised in the Final Rejection dated
November 5, 2003 remaining for resolution is are Claims 1-20 on
appeal patentable, under 35 U.S.C. § 103(a), in light of the
combination of Inagaki and Pavlovic.

### X.  THE REFERENCES

Inagaki discloses a video conferencing system that
detects the voice of a speaking attendee and highlights the PIP
of the speaking attendee to distinguish the same from the other
attendees.  The system of Inagaki does not recognize and/or

-9-

respond to spoken commands.  Pavlovic is directed to recognition of spoken commands.

### XI.  <u>GROUPING OF THE CLAIMS</u>

The prior art rejections of issue herein apply to more than one claim.  However, Appellant submits that at least dependent claims 2, 5, and 6 patentably distinguish over the cited references independently of their base claim.

### XII.  <u>APPELLANT'S ARGUMENTS</u>

The rejection of Claims 1-20, on appeal, under 35 U.S.C. § 103(a), as unpatentable over Inagaki in view of Pavlovic is improper.

Claim 1 recites:

"a processor … configured … to change a PIP display characteristic in response to a received audio command and a related gesture from a user."

Independent claims 11, 15, 19 and 20 contain similar recitations regarding an audio command or indication and a related gesture.

Inagaki discloses a video conferencing system that **<u>detects the voice of a speaking attendee</u>** and **<u>highlights</u>** the PIP of the speaking attendee to distinguish the same from the other attendees.  Therefore, Inagaki does not disclose or suggest "a processor … configured … to change a PIP display characteristic in response to a received audio **<u>command</u>** and a related gesture from a user."  **<u>Inagaki merely detects the presence of speech.</u>**

-10-

**Inagaki does not teach the detection of a content of the speech to determine if a command is being spoken.** Therefore, as discussed more fully below, since the system of Inagaki does not recognize and/or respond to spoken commands, there is no motivation or suggestion to combine it with the system of Pavlovic that is directed to recognition of spoken commands.

Recently the U.S. Court of Appeals for the Federal Circuit (the "Federal Circuit") restated the legal test applicable to rejections under 35 U.S.C. § 103(a) (*In re Rouffet*, 47 USPQ2d 1453 (Fed. Cir., July 15, 1998)). The Court stated:

> [V]irtually all [inventions] are combinations of old elements. Therefore an Examiner may often find every element of a claimed invention in the prior art. Furthermore, rejecting patents solely by finding prior art corollaries for the claimed elements would permit an Examiner to use the claimed invention itself as a blueprint for piecing together elements in the prior art to defeat the patentability of the claimed invention. Such an approach would be "an illogical and inappropriate process by which to determine patentability." To prevent the use of hindsight based on the invention to defeat patentability of the invention, this court requires the Examiner to show a motivation to combine the references that create the case of obviousness. The Board [of Appeals] did not, however, explain what specific understanding or technological principle within the knowledge of one of ordinary skill in the art would have suggested the combination. …To counter this potential weakness in the obviousness construct **the suggestion to combine requirements stands as a critical**

**safeguard against hindsight analysis and rote
application of the legal test for obviousness.**

*In re Rouffet*, 47 USPQ2d 1457-58 (Fed. Cir.,
July 15, 1998) (citations omitted, emphasis
added).

More recently, the Federal Circuit again dealt with

what is required to show a motivation to combine references

under 35 U.S.C. § 103(a). In this case the court reversed the

decision of the Board of appeals stating:

> [R]ather then pointing to specific information
> in Holiday or Shapiro that suggest the
> combination..., the Board instead described in
> detail the similarities between the Holiday and
> Shapiro references and the claimed invention,
> noting that one reference or the other-in
> combination with each other... described all of
> the limitations of the pending claims. Nowhere
> does the Board particularly identify any
> suggestion, teaching, or motivation to combine
> the ... references, nor does the Board make
> specific-or even inferential-findings concerning
> the identification of the relevant art, the
> level of ordinary skill in the art, the nature
> of the problem to be solved, or any factual
> findings that might serve to support a proper
> obviousness analysis.
>
> *In re Dembiczak*, 50 USPQ2d 1614, 1618 (Fed.
> Cir., April 28, 1999) (citations omitted).

Thus, from both *In re Rouffet* and *In re Dembiczak* it

is clear that the Federal Circuit requires a specific

identification of a suggestion, motivation, or teaching why one

of ordinary skill in the art would have been motivated to

select the references and combine them. This the Examiner has

not done. The Examiner only states that it would be obvious

-12-

"to one of ordinary skill in the art at the time of the invention was made to use a 'received audio command and a related gesture from a user' as taught by Pavlovic in the apparatus of Inagaki, because of the motivation provided by Pavlovic "Psychological studies, for example, show that <u>people prefer to use hand gestures in combination with speech in a virtual environment</u>, since they allow the user to interact without special training or special apparatus'." (see page 3 of the Final Official Action, emphasis in original).

However, Applicant respectfully submits that the motivation argued by the Examiner could have only been gleaned from the present application. As discussed above, Inagaki neither discloses nor suggests the use of either a hand signal or audio command, therefore, those of ordinary skill in the art would not been suggested or motivated to combine its teachings with those of Pavlovic.

Thus, Applicant respectfully submits that the Examiner, without identifying a suggestion, motivation, or teaching for combining the references, has used impermissible hindsight to reject claims 1-20 under 35 U.S.C. 103(a). As discussed above, the Federal Circuit in *In re Rouffet* stated that virtually all inventions are combinations of old elements. Therefore an Examiner may often find every element of a claimed invention in the prior art. To prevent the use of hindsight

-13-

based on the invention to defeat patentability of the invention, the Examiner is required to show a motivation to combine the references that create the case of obviousness. Applicant respectfully submits that the Examiner has not met this burden.

In light of the state of the law as set forth by the Federal Circuit and the Examiner's lack of specificity with regard to the motivation to combine the cited references, the applicant respectfully submits that the rejections for obviousness under 35 U.S.C. § 103(a) lack the requisite motivation and must be withdrawn.

Therefore, the Applicant respectfully submits that the rejection of claims 1-20 for obviousness under 35 U.S.C. § 103(a) lacks the requisite motivation and must be withdrawn.

Furthermore, with regard to independent claim 11, the same recites:

> "determining whether the received audio indication is one of a plurality of expected audio indications:
> analyzing a gesture of the user if the received audio indication is one of the plurality of expected audio indications"

The Applicant respectfully submits that neither of the cited references teach or suggest the recited features of claim 11. As discussed above, Inagaki merely teaches a voice detection to determine whether an attendee is speaking. Thus, Inagaki does not teach or suggest "determining whether the

-14-

received audio indication is one of a plurality of expected

audio indications." Pavlovic teaches the user issuing a spoken

command and gesture simultaneously (page 123, right column)

because studies show speech and gestures are preferred by some

people and because they are produced by the same mental

concept. Page 123 of Pavlovic states:

> "...the user utters the command 'select
> this', while **simultaneously** pointing at an
> object. To perform an object, the user utters
> a command such as 'move left' while
> **simultaneously** performing the 'move left'
> gestural action." (emphasis added).

Thus, Pavlovic does not teach or suggest "analyzing a

gesture of the user if the received audio indication is one of

the plurality of expected audio indications." Claim 20 contains

similar recitations and patentably distinguishes over the cited

references for at least the same reasons as set forth above

with regard to claim 11.

Still further, with regard to claim 15, the same

contains program segments for performing similar steps as

recited in claim 11 and patentably distinguishes over the cited

references for at least the same reasons as set forth above

with regard to claim 11.

Therefore, assuming arguendo that the combination of

Inagaki and Pavlovic is proper (which Applicant maintains it is

not), at least independent claims 11, 15, and 20 are not

rendered obvious by the cited references because neither the

-15-

Inagaki patent nor the Pavlovic publication, whether taken alone or in combination, teach or suggest a method having the features recited therein and discussed above. Accordingly, claims 11, 15, and 20 patentably distinguish over the prior art and are allowable. Claims 12-18 being dependent upon claims 11 and 15 are thus allowable therewith.

Lastly, Applicant respectfully submits that at least dependent claims 2, 5, and 6 patentably distinguish over the cited references independently of their base claim (claim 1) and are allowable.

With regard to claim 2, the same recites "the PIP display characteristic is at least one of a position of the PIP on the display and a display size of the PIP." As discussed above, Inagaki merely discloses a video conferencing system that detects the voice of a speaking attendee and **highlights** the PIP of the speaking attendee to distinguish the same from the other attendees. In the Final Official Action, the Examiner cites Figure 8a of Inagaki as showing the features of claim 2. However, Figure 8a merely shows highlighting the PIP display of speaker B corresponding to the speaker (B) in the main display. Thus, Inagaki does not change "at least one of a position of the PIP on the display and a display size of the PIP" as is recited in claim 2.

With regard to claims 5 and 6, claim 5 recites

-16-

"the processor is configured to analyze image information received from the user after the audio indication is received to identify the change in the PIP display characteristic that is expressed by the received gesture."

In the Final Official Action, the Examiner argues that Figures 8a and 8b of Inagaki and Figures 6-8 of Pavlovic show the features of claims 5 and 6. However, as discussed above with regard to claim 2, Figures 8a and 8b merely show highlighting a PIP display corresponding to the speaker in the main display (speaker B in Figure 8a and speaker D in Figure 8B). Furthermore, the description on page 123 of Pavlovic discussed above with regard to claim 11 accompanies Figures 6-8 cited by the Examiner. Thus, as concluded above, Pavlovic teaches the user issuing a spoken command and gesture **simultaneously**.
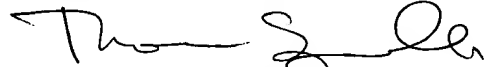
Therefore, at least claims 2, 5, and 6, patentably distinguish over the cited references independently of their base claim (1) and are allowable.

### XIII. CONCLUSION

In view of the remarks submitted hereinabove, the references applied against Claims 1-20 on appeal do not render those claims unpatentable under 35 U.S.C. § 103(a). Thus, Appellant submits that the § 103(a) rejection is in error and must be reversed.

-17-

The Commissioner is hereby authorized to charge any additional fees or credit any overpayment in connection herewith to Deposit Account No. 19-1013/SSMP. A triplicate copy of this sheet is enclosed.

Respectfully submitted,

Thomas Spinelli
Registration. No.: 39,533

SCULLY SCOTT MURPHY & PRESSER
400 Garden City Plaza
Garden City, New York  11530
(516) 742-4343
TS/cm

APPENDIX

CLAIMS ON APPEAL:   CLAIMS 1-20
Application Serial No. 09/896,199

1.   (Previously Presented) A video display device comprising:

a display configured to display a primary image and a picture-in-picture image (PIP) overlaying the primary image; and

a processor operatively coupled to the display and configured to receive a first video data stream for the primary image, to receive a second video data stream for the PIP, and to  change a PIP display characteristic in response to a received audio command and a related gesture from a user.

2.   (Original) The video display device of Claim 1, wherein the PIP display characteristic is at least one of a position of the PIP on the display and a display size of the PIP.

3.   (Original) The video display device of Claim 1, comprising:

a microphone for receiving the audio indication from the user; and

a camera for acquiring an image of the user containing the related gesture.

-19-

4. (Original) The video display device of Claim 1, wherein the processor is configured to analyze audio information received from the user to identify when a PIP related audio indication is intended by the user.

5. (Original) The video display device of Claim 1, wherein the processor is configured to analyze image information received from the user after the audio indication is received to identify the change in the PIP display characteristic that is expressed by the received gesture.

6. (Original) The video display device of Claim 5, wherein the image information is contained in a sequence of images and wherein the processor is configured to analyze the sequence of images to determine the received gesture.

7. (Original) The video display device of Claim 1, wherein the image information is contained in a sequence of images and wherein the processor is configured to determine the received gesture by analyzing the sequence of images and determining a trajectory of a hand of the user.

8. (Original) The video display device of Claim 1, wherein the processor is configured to determine the received gesture by analyzing an image of the user and determining a posture of a hand of the user.

-20-

9.    (Original) The video display device of Claim 1, wherein the video display device is a television.

10.    (Original) The video display device of Claim 1, wherein the image is a sequence of images of the user containing the user gesture, the video display device comprising a camera for acquiring the sequence of images of the user.

11.    (Original) A method of controlling a display characteristic of a picture-in-picture display (PIP) overlaying a primary display, the method comprising:

receiving an audio indication from a user;

determining whether the received audio indication is one of a plurality of expected audio indications;

analyzing a gesture of the user if the received audio indication is one of the plurality of expected audio indications; and

controlling the display characteristic if the gesture is a gesture related to the received audio indication.

12.    (Original) The method of Claim 11, wherein analyzing the gesture comprises:

receiving a sequence of images; and

analyzing the sequence of images to determine the gesture.

-21-

13.  (Original) The method of Claim 11, wherein analyzing the gesture comprises:

receiving a sequence of images;

analyzing the sequence of images to determine a trajectory of a hand of the user; and

determining the gesture by the determined trajectory.

14.  (Original) The method of Claim 11, wherein analyzing the gesture comprises:

analyzing an image of the user to determine a posture of a hand of the user; and

determining the gesture by the determined posture.

15.  (Original) A program segment stored on a processor readable medium for controlling a display characteristic of a picture-in-picture display (PIP) overlaying a primary display, the program segment comprising:

a program segment for controlling receipt of an audio indication;

a program segment for determining whether a received audio indication is one of a plurality of stored audio indications;

a program segment for analyzing a gesture of the user if the received audio indication is one of the plurality of stored audio indications; and

a program segment for controlling the display

-22-

characteristic if the gesture is a gesture related to the received audio indication.

16. (Original) The program segment of Claim 15, wherein the program segment for analyzing the gesture comprises:

a program segment for controlling receipt of a sequence of images; and

a program segment for analyzing the sequence of images to determine the gesture.

17. (Original) The program segment of Claim 15, wherein the program segment for analyzing the gesture comprises:

a program segment for controlling receipt of a sequence of images;

a program segment for analyzing the sequence of images to determine a trajectory of a hand of the user; and

a program segment for determining the gesture by the determined trajectory.

18. (Original) The program segment of Claim 15, wherein the program segment for analyzing the gesture comprises:

a program segment for analyzing an image of the user to determine a posture of a hand of the user; and

-23-

a program segment for determining the gesture by the determined posture.

19.    (Previously Presented) A video display device comprising:

a display configured to display a primary image and a picture-in-picture image (PIP); and

a processor operatively coupled to the display and configured to receive a first video data stream for the primary image, to receive a second video data stream for the PIP, and to change a PIP display characteristic in response to a received audio indication and a related gesture from a user, wherein the PIP display characteristic is at least one of a position of the PIP on the display and a display size of the PIP.

20.    (Previously Presented) A video display device comprising:

a display configured to display a primary image and a picture-in-picture image (PIP); and

a processor operatively coupled to the display and configured to receive a first video data stream for the primary image, to receive a second video data stream for the PIP, and to  change a PIP display characteristic in response to a received audio indication and a related gesture from a user, wherein the processor is configured to analyze image

-24-

information received from the user after the audio indication

is received to identify the change in the PIP display

characteristic that is expressed by the received gesture.